fGlobal phylogeography and species delimitation in the "sexy shrimp" *Thor amboinensis*

Undergraduate Research Thesis

Presented in Partial Fulfillment of the Requirements for graduation "with Honors Research Distinction" in the undergraduate colleges of The Ohio State University

by

Natalie Hamilton

The Ohio State University
May 2015

Projection Advisor: Professor Marymegan Daly, Department of EEOB

Abstract

Marine species that exist in continuously distributed populations and disperse *via* pelagic larvae are expected to have significant gene flow between populations. However, physical barriers to dispersal and isolated peripheral habitats may lead to highly structured populations and speciation events. For species with cosmopolitan distributions, these barriers may limit dispersal and gene flow. In order to explore the relationship between gene flow, dispersal potential, and possible speciation events we used mtDNA sequences from cytochrome c oxidase subunit I (COI) from the 'sexy' shrimp *Thor amboinensis,* a circumtropical symbiotic species found on coral reefs. We assessed the phylogeography of *T. amboinensis* across an east-west distribution in the Caribbean, and investigated the potential for cryptic speciation between Caribbean and Indo-Pacific populations. Pair-wise ϕ_{ST} values indicate no significant genetic structuring between Bermuda, Florida Keys, Honduras, and Mexico. Haplotypes were shared across all sample localities in the Caribbean, suggesting panmixia between the regions of the Caribbean and long distance dispersal potential. However, divergent un-linked haplotypes were found between the Caribbean and the Indo-Pacific, evidence for the existence of cryptic species. Pairwise sequence divergence between Caribbean and Indo-Pacific populations was 14%, far exceeding the commonly used 3-5% cutoff used to delimit crustacean species with COI barcodes (Hebert *et al*. 2003; Barrett & Hebert, 2005; Kress & Erickson, 2008). The species delimitation program Bayesian General Mixed Yule Coalescent (bGMYC) showed that there were two separate species groups at the 95% confidence level. Although there is evidence that long distance dispersal is a characteristic of Caribbean populations, the Caribbean and Indo-Pacific populations are highly diverged and *T. amboinensis* likely represents at least two species. As this species is highly valued in the ornamental aquarium trade, recognizing that *T. amboinensis* is

actually a cryptic species complex will be helpful for scientifically based collecting protocols and conservation initiatives.

Introduction

Species are the lowest level of taxonomic classification, and proper species identification and delimitation is a crucial first step for ecological and evolutionary research (De Queiroz, 2007; Pante *et al.* 2015; Satler *et al.* 2013). While species concepts and delimitation has been highly contended and a number of competing concepts exist, the common goal of species delimitation is to identify independently evolving metapopulations (de Quieroz 2005, 2007). In evolutionary biology, proper species delimitation and the identification of cryptic species is important to avoid biased estimates of genetic connectivity, migration, and gene flow between populations (Pante *et al*. 2015).

For cosmopolitan or circumtropical species, taxonomic status as a single species implies some level of gene flow and migration between populations. However many habitats are not continuous and major barriers to gene flow and migration can exist, leading to cryptic speciation events that are difficult to detect without in-depth taxonomic investigation. In marine ecosystems, taxonomic resources for many taxa are limited, leading to poor species descriptions and estimates of species boundaries (reviewed by Pante *et al*. 2015). As a result, genetic surveys using DNA barcodes are often an initial first step towards species discovery and delimitation (Hebert *et al*. 2003; Barrett & Hebert, 2005; Kress & Erickson, 2008). Marine crustaceans are a historically diverse group morphologically and the advent of DNA barcoding has lead to the discovery of even higher biodiversity within the group (Costa *et al*. 2011). However, tropical reef habitat is not continuous. Many reef ecosystems are separated by vast expanses of open

ocean, and approximately 4.5 million years ago, the Isthmus of Panama appeared between North and South America (Knowlton & Weight 1998), effectively separating the Atlantic and Pacific Oceans, which up until this point had been able to mix freely. The formation of the Isthmus created a physical barrier to dispersal of marine species, and investigation of cryptic species between the two populations could explore the idea of alternative dispersal routes between populations. Thus, Caribbean populations may be sufficiently isolated from the Indo-Pacific to be considered a separate species. If there is evidence of cryptic speciation in the two regions, a time line of speciation could be examined.

The "sexy shrimp" *Thor amboinensis* (Figure 1) is a circumtropical sea anemone symbiont that is currently described as a single species ranging from the Red Sea through the Indo-Pacific, Caribbean, and Atlantic terminating at the Canary Islands off the coast of Africa. This broad range coupled with the high incidence of cryptic diversity in crustaceans makes it a good candidate to investigate species delimitation in marine crustaceans. In their larval stage, *T. amboinensis* are suspended in the water column for 35-40 days (Sarver, 1979). The length of their pelagic larval stage means that these shrimp have a great opportunity for dispersal. This shrimp is harvested in tropical waters worldwide. It is currently described as a single species, but given its global geographic range, *T. amboinensis* may actually represent multiple cryptic lineages. Given these characteristics, *T. amboinensis* represents a potentially important model species to understand the evolution of globally distributed tropical marine invertebrates.

Fig 1. Representative image of *Thor amboinensis*. source: Todoreef.com

We investigated the phylogeography and species delimitation of *T. amboinensis* populations in the Caribbean (Florida, Honduras, Bermuda, Mexico, and St. Thomas) and the Indo-Pacific (Philippines and Bali) through sequencing of the mitochondrial gene cytochrome oxidase subunit 1 (COI). We show that while dispersal ability is apparently high in the Caribbean, these populations are highly divergent from the Indo-Pacific and likely represent cryptic species. Through examining the diversity of *T. amboinensis* populations we increase our understanding of the speciation processes in tropical marine habitats, and contribute to the understanding and documentation of biodiversity in these ecosystems. Discovering and documenting biodiversity is important as extinction rates have increased over the past few decades (Pimm *et al*. 1995), especially for commercially harvested animals with little taxonomic or ecological information.

Methods

A total of 172 *Thor amboinensis* samples were collected from 11 coral reef sites (Figure 2). These sites include samples from Florida (Fort Lauderdale, Upper Keys, Middle Keys, and Lower Keys), Honduras (Utila, Cayos Cochinos), Mexico, Bermuda, St. Thomas, Bali, and the Philippines. The samples from Bali were purchased through the aquarium trade business Reef Systems Coral Farm, Inc. in New Albany, OH. Other samples were collected via SCUBA by PhD student Ben Titus, preserved in 100% EOH or RNAlater in the field, and transferred to The Ohio State University.

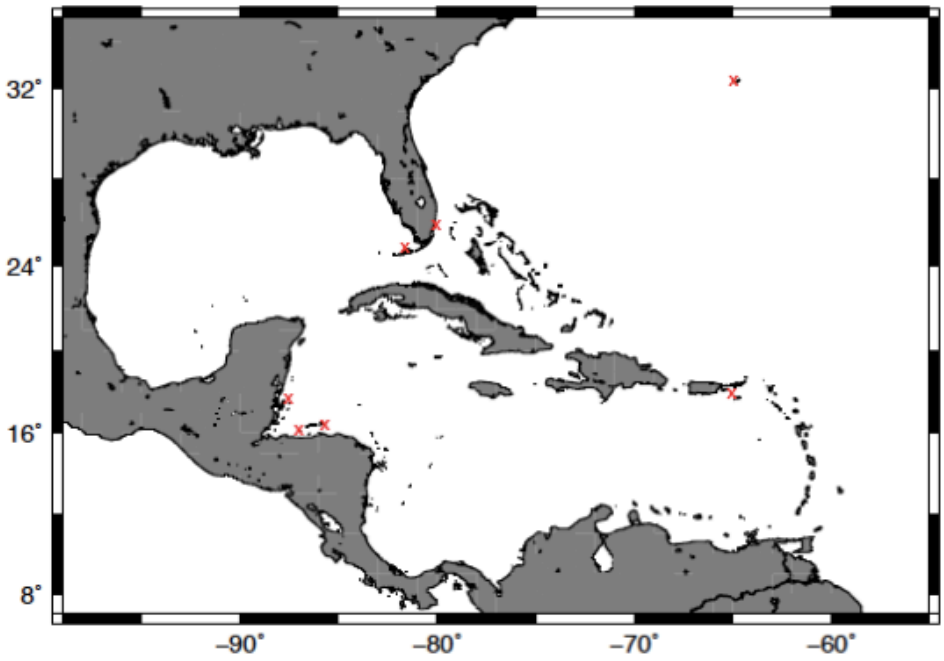

Figure 2: Range map of where *Thor amboinensis* were sampled in Caribbean. Additional samples were collected from Bali and the Philippines.

Total genomic DNA was isolated using DNeasy Blood and Tissue Kits (QIAGEN Inc.) and stored at -20C. Using universal primers, a fragment of the mtDNA cytochrome oxidase subunit 1 (COI), a ~650 bp fragment was amplified using polymerase chain reaction (PCR). COI

has been shown as a reliable bioindentification system for animals (Hebert *et al*. 2003). These reactions were carried out in 25-ul volumes using IllustraTM puReTaqTM Ready-To-Gotm PCR beads (GE Healthcare). PCRs were run in an epigradient Mastercycler (Eppendorf) with run conditions following Santos (2006). The amplified samples were sequenced in both directions at Beckman Coulter Genomics. Consensus sequences were created using SEQUENCHER 4.9 (Gene Codes Corp., Ann Arbor, MI, USA). Consensus sequences were aligned using Muscle in MEGA v 6.06 (Kumar *et al*. 2008). The low confidence reads were trimmed from both ends of the aligned consensus sequences, resulting in a final sequence length of 572 bp.

Nucleotide (π) and haplotype (h) diversity indices (Table 1) were calculated according to Nei (1987) using DNASP v 5.1.0 (Librado & Rozas 2009). Pairwise distances were calculated within and between haplotypes using Mega v 6.06. To assess population structure and genetic variances among sampling locations, pairwise ϕst were calculated using ARLEQUIN v 3.11 (Excoffier *et al*.2005). An analysis of molecular variance (AMOVA) was conducted using ARLEQUIN to assess how genetic diversity is partitioned hierarchically; within populations (ϕst), among populations within groups (ϕsc), and among groups (ϕct). Demographic patterns were assessed using Tajima's D (Tajima 1989) and Fu's F neutrality tests using ARLEQUIN (Table 1). Tajima's D and Fu's F determine whether variations in sequence data fit the neutral model and can offer valuable information regarding the demographic history of the population (Rand, 2006).

In order to understand the species delimitation, statistical parsimony networks were visualized using TCS v 1.21 with the default settings to create parsimonious branch connections at the 95% confidence level (Clement *et al.* 2000) (Figure 3). These results were compared to pair-wise distances computed in MEGA v. 6.06. BGMYC program was run as described in (Reid

& Carstens, 2012) in order to further visualize the relationships between the *T. amboinensis* populations (Figure 4).

Table 1: Genetic diversity indices and neutrality tests of *T. amboinensis*

Population	Site	n	nh	π	h	Fu's F	Tajima's D
Indo-Pacific							
	Bali	9	7	0.00728	0.971	-1.67243	-.61495
	Philippines	2	2	0.00524	1	1.09861	0
Caribbean							
	Bermuda	28	5	0.00062	0.27	-2.83419*	-1.95558*
	Honduras	26	7	0.00093	0.471	-4.17968*	-1.80901*
	Mexico	29	6	0.00071	0.374	-3.55063*	-1.71434*
	Florida	50	9	0.00056	0.297	-10.95876*	-2.23984*

n, number of samples; nh, number of unique haplotypes; π, nucleotide diversity; h, haplotype diversity
*$p<0.05$

Results

The consensus sequences were trimmed and a total of 572 bp of COI was obtained for the 172 samples of *Thor amboinensis*. There were no insertions, deletions, or stop codons. In total, 32 unique haplotypes were recovered (23 Caribbean; 9 Indo-Pacific), and 96 of the sites were polymorphic. Nucleotide (π) and haplotype (h) diversity were higher in the Indo-Pacific locals than in the Caribbean (Table 1). Neutrality tests (Tajima's D and Fu's F; Table 1) were significantly negative in the Caribbean populations, while the values for the Indo-Pacific populations were not significant.

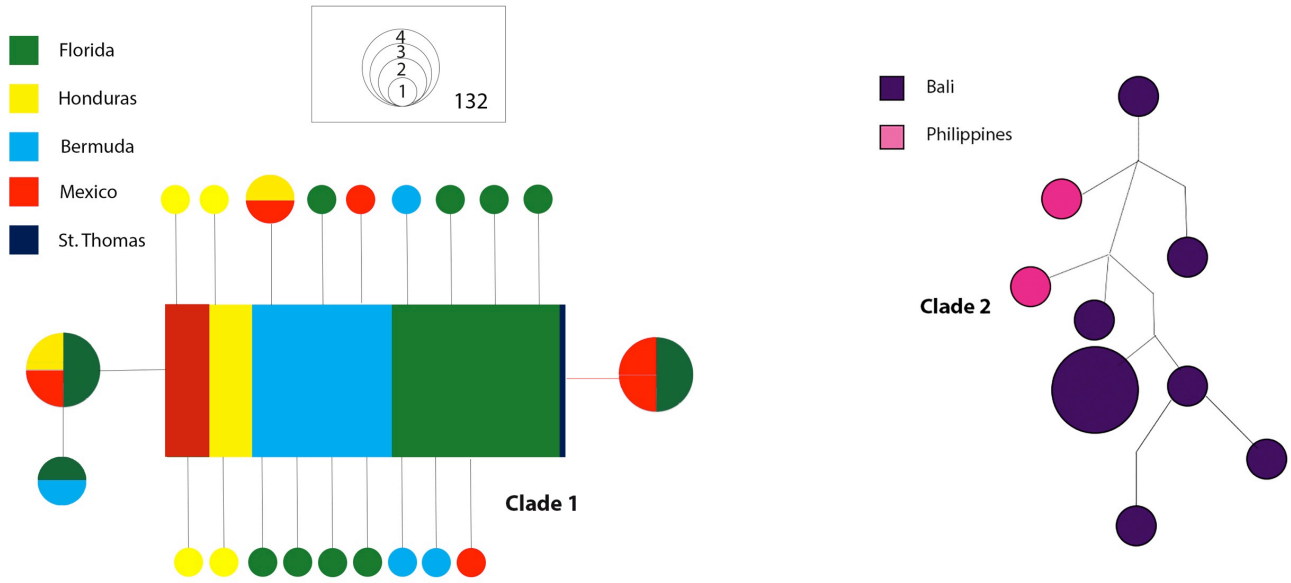

Figure 3: Statistical parsimony haplotype network for *Thor amboinensis*. Haplotype color represents sampling location. Rectangle represents inferred ancestral haplotype groups and circles represent smaller haplotype samples. Rectangle and circle size represent relative frequency with which an individual haplotype was sampled.Each branch represents a single mutation step. Larger (left) network = Clade 1; smaller (right) network= Clade 2

Pair-wise distance within each clade was 0% (Caribbean) and 0% - 1 % (Indo-Pacific), while pairwise distance between Caribbean and Indo-Pacific clades was 14%. Population genetic structure analysis returned significant ϕst values between Indo-Pacific and Caribbean regions. The AMOVA of *T. amboinensis* populations using all samples found that a significant proportion of the genetic variance was found within populations (85.11% $p<0.0001$). The proportion of genetic variation found among populations within groups (2.03%; $p=0.356$) and among groups (12.86%; $p=0.051$) was not significant. However, the values for among groups is approaching significance. An AMOVA with Caribbean samples only reveals highly insignificant values for among group genetic variation, so the difference between the Indo-Pacific and the Caribbean groups is what is driving the genetic variation among-groups.

Statistical parsimony analysis (TCS) yielded two separate haplotype networks (Fig. 3) at the 95% confidence level. The larger haplotype network (henceforth known as Clade 1) is made of 161 samples and encompasses all of the individuals from the Caribbean range. The second, smaller clade (henceforth known as Clade 2) consists of 11 samples from the Indo-Pacific.

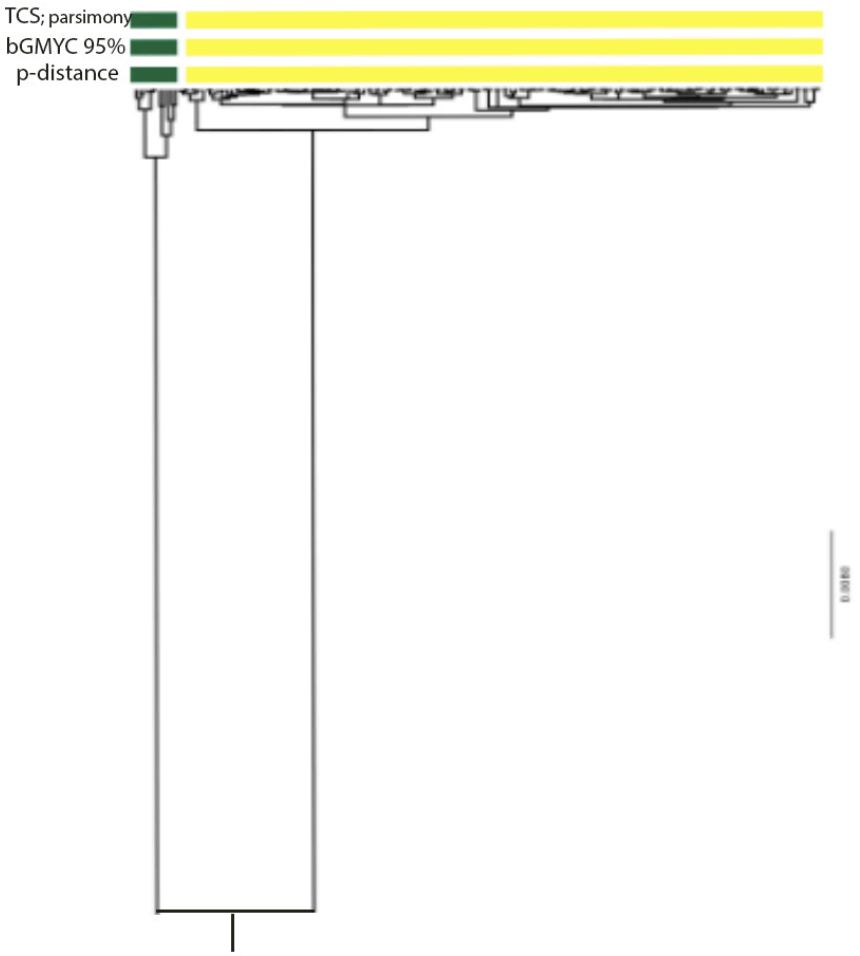

Figure 4: Three species delimitation models. Yellow represents Clade 1 and Green represents Clade 2. P-distance values of 13-14% between the two clades.

Pair-wise distance analysis showed a 13-14% distance between Clade 1 and Clade 2, which far exceeds the 3-5% threshold for speciation. BGMYC analysis further supports these results; two distinct species groupings were produced at the 95% confidence level (Figure 4).

Discussion:

Genetic diversity indices nucleotide (π) and haplotype (h) suggest that genetic diversity is greatest in the Philippines and Bali. This is to be expected, as the Indo-Pacific is the highest site of biodiversity for the genus *Thor* (De Grave *et al.* 2014). The lower genetic diversity in the Caribbean region can be explained by the population genetics theory, which predicts that genetic diversity steadily decreases along the species range as a result of consecutive bottleneck effects (Excoffier *et al.* 2009). The neutrality tests (Fu's F and Tajima's D) have values that further support the recent population expansion of the Caribbean locations. The significantly negative values for both neutrality tests suggest that the *T. amboinensis* populations in the Caribbean were founded by a population of small size. The ϕst values of the Caribbean populations indicated low population structure in the Caribbean region. This means that there is significant gene flow between the populations in the Caribbean. There are shared haplotypes between all Caribbean populations, which suggests long distance dispersal potential, which is to be expected due to the extended larval dispersal stage of *T. amboinensis*.

The distinctness of the two clades and a pair-wise analysis give evidence that the Caribbean and Indo-Pacific populations have been separated for 6.09-8.24 million years. This is estimated based on the current estimation of COI mutation rates in arthropods to be at 1.7%-2.3% for every million years of separation (Bower, 1994). Crustaceans tend to hold a high level of biodiversity and species delimitation studies using COI require a 3-5% K2P pairwise

nucleotide difference to support the existence of cryptic species (Plaisance *et al*. 2009). K2P distance between clades was ~14%, which exceeds the minimum threshold for species. The bGMYC analysis created two separate species groupings at the 95% confidence level. Finally, all three measures of species delimitation suggest that Indo-Pacific and Caribbean populations are actually represent two separate species. Identification and documentation of cryptic species is important in the understanding of marine populations as opened or closed systems. The delimitation of cryptic species is an important basis for ecological studies. By understanding these species as two separately evolving metapopulations, we can avoid connectivity biases (Reviewed by Pante *et al.* 2015). Cryptic speciation also has the potential to impact collection practices.

Although the Isthmus of Panama creates a physical barrier to gene dispersal, it shut off the flow between the Atlantic and the Pacific Oceans around 4.5 million years ago and our results have shown that the populations of *T. amboinensis* have been genetically isolated for a longer period of time (6-8 my). As the gene flow between the Caribbean populations is evidence of long range of *T. amboinensis* larval dispersal, it is possible there are other physical barriers to gene flow between the Caribbean and Indo-Pacific populations. Increased sampling and DNA sequencing from throughout the Indo-Pacific range is needed in order to fully understand the extent of genetic diversity in *Thor amboinensis*. Future morphological research (conducted by associates at Clemsen University) will further verify these two populations as separate species. At minimum, COI genetic sequencing suggests two separate taxa.. This research refines our understanding of speciation and species boundaries in circumtropical marine invertebrates and will be important for conservation applications. As the species in the Caribbean and the Indo-Pacific seem to have no gene flow, oversampling in one region could cause one cryptic species

to become endangered and have no access to repopulation by dispersal from the other populations.

References:

Barrett, R. D., & Hebert, P. D. (2005). Identifying spiders through DNA barcodes. *Canadian Journal of Zoology*, *83*(3), 481-491.

Brower, A. V. (1994). Rapid morphological radiation and convergence among races of the butterfly Heliconius erato inferred from patterns of mitochondrial DNA evolution. *Proceedings of the National Academy of Sciences*, *91*(14), 6491-6495.

Clement, M., Posada, D. C. K. A., & Crandall, K. A. (2000). TCS: a computer program to estimate gene genealogies. *Molecular ecology*, *9*(10), 1657-1659.

Costa, F. O., DeWaard, J. R., Boutillier, J., Ratnasingham, S., Dooh, R. T., Hajibabaei, M., & Hebert, P. D. (2007). Biological identifications through DNA barcodes: the case of the Crustacea. *Canadian Journal of Fisheries and Aquatic Sciences*, *64*(2), 272-295.

De Grave, S., Li, C. P., Tsang, L. M., Chu, K. H., & Chan, T. Y. (2014). Unweaving hippolytoid systematics (Crustacea, Decapoda, Hippolytidae): resurrection of several families. *Zoologica Scripta*, *43*(5), 496-507.

De Queiroz, K. (2005). Ernst Mayr and the modern concept of species. *Proceedings of the National Academy of Sciences*, *102*(suppl 1), 6600-6607

De Queiroz, K. (2007). Species concepts and species delimitation. *Systematic biology*, *56*(6), 879-886.

Excoffier, L., Laval, G., & Schneider, S. (2005). Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evolutionary bioinformatics online*, *1*, 47.

Hebert, P. D., Ratnasingham, S., & de Waard, J. R. (2003). Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society of London B: Biological Sciences*, *270*(Suppl 1), S96-S99.

Hoekstra, J. M., Boucher, T. M., Ricketts, T. H., & Roberts, C. (2005). Confronting a biome crisis: global disparities of habitat loss and protection. *Ecology Letters*, *8*(1), 23-29.

Knowlton, N., & Weigt, L. A. (1998). New dates and new rates for divergence across the Isthmus of Panama. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *265*(1412), 2257-2263.

Kumar, S., Nei, M., Dudley, J., & Tamura, K. (2008). MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. *Briefings in bioinformatics*, *9*(4), 299-306.

Librado, P., & Rozas, J. (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, *25*(11), 1451-1452.

Pante, E., Puillandre, N., Viricel, A., Arnaud-Haond, S., Aurelle, D., Castelin, M., Chenuil, A., Destombe, C., Forcioli, D., Valero, M., Viard, F. and Samadi, S. (2015), Species are hypotheses: avoid connectivity assessments based on pillars of sand. *Molecular Ecology,* 24: 525–544. doi: 10.1111/mec.13048

Plaisance, L., Knowlton, N., Paulay, G., & Meyer, C. (2009). Reef-associated crustacean fauna: biodiversity estimates using semi-quantitative sampling and DNA barcoding. *Coral Reefs*, *28*(4), 977-986.

Rand, D. M. (1996). Neutrality tests of molecular markers and the connection between DNA polymorphism, demography, and conservation biology. *Conservation Biology*, 665-671.

Rozas, J., Sánchez-DelBarrio, J. C., Messeguer, X., & Rozas, R. (2003). DnaSP, DNA

polymorphism analyses by the coalescent and other methods. *Bioinformatics*, *19*(18), 2496-2497.

Santos, S. R. (2006). Patterns of genetic connectivity among anchialine habitats: a case study of the endemic Hawaiian shrimp *Halocaridina rubra* on the island of Hawaii. *Molecular Ecology*, *15*(10), 2699-2718.

Sarver, D. (1979). Larval culture of the shrimp Thor amboinensis (De Man, 1888) with reference to its symbiosis with the anemone *Antheopsis papillosa* (Kwietniewski, 1898). *Crustaceana. Supplement*, 176-178.

Satler, J. D., Carstens, B. C., & Hedin, M. (2013). Multilocus species delimitation in a complex of morphologically conserved trapdoor spiders (Mygalomorphae, Antrodiaetidae, Aliatypus). *Systematic Biology*, *62*(6), 805-823.

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, *123*(3), 585-595.

Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic acids research*, *22*(22), 4673-4680.

Witt, J. D., Threloff, D. L., & Hebert, P. D. (2006). DNA barcoding reveals extraordinary cryptic diversity in an amphipod genus: implications for desert spring conservation. *Molecular Ecology*, *15*(10), 3073-3082.